PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/81, C12P 21/02

A1

(11) International Publication Number: **WO 00/20610**

(43) International Publication Date: 13 April 2000 (13.04.00)

(21) International Application Number: PCT/US99/23351

(22) International Filing Date: 7 October 1999 (07.10.99)

(30) Priority Data:
60/103,414 7 October 1998 (07.10.98) US

(71) Applicant: THE CHILDREN'S MEDICAL CENTER CORPORATION [US/US]; 300 Longwood Avenue, Boston, MA 02115 (US).

(72) Inventors: BOEHM, Thomas; 15 Greenstrat #1, Brookline, MA 02146 (US). FOLKMAN, M., Judah; 18 Chatham Circle, Brookline, MA 02146 (US).

(74) Agents: WARREN, William, L. et al.; Jones & Askew, LLP, 2400 Monarch Tower, 3424 Peachtree Road, Atlanta, GA 30326 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: DISRUPTION OF THE KEX1 GENE IN *PICHIA* AND METHODS OF FULL LENGTH PROTEIN EXPRESSION

(57) Abstract

Methods and compositions for recombinant expression of full length proteins in *Pichia*. The invention provides genetic constructs containing a disruption in the KEX1 gene that prevent cleavage of basic amino acids, such as lysine, from the carboxy terminal of proteins expressed therewith.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

DISRUPTION OF THE KEX1 GENE IN *PICHIA* AND METHODS OF FULL LENGTH PROTEIN EXPRESSION

Statement of Governmental Rights

This work was supported by an NIH grant. The U.S. Government may have certain rights in this invention.

Related Applications

This application claims priority to U.S. Provisional Patent Application Serial No. 60/103,414 filed October 7, 1998.

Technical Field

The invention relates to protein expression and more specifically relates to genetically modified yeasts belonging to the genus *Pichia* and to their use to produce recombinant proteins.

Background of the Invention

It is possible to modify microorganisms in order to make them produce proteins of interest such as, for example, mammalian proteins, artificial proteins, chimeric proteins, and the like. In particular, numerous genetic studies have been performed on the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae*. More recently, genetic tools have been developed so as to use the yeast *Pichia pastoris* as host cell for the production of recombinant proteins.

Gram to kilogram quantities of proteins must be

recombinantly produced for clinical trials. If such trials show promising results, hundreds or thousands of kilograms are typically necessary to meet the demands of patients worldwide.

Expression in *Pichia* offers many of the benefits of *E. coli* (high-level expression, easy scale-up, and inexpensive growth) combined with many of the advantages of expression in an eukaryotic system (protein processing, folding, and posttranslational modifications).

P. pastoris is a methylotrophic yeast. In the absence of a repressing carbon source, such as glucose, *P. pastoris* is able to use a metabolic pathway that allows it to utilize methanol as a carbon source. The alcohol oxidase promoter (AOX1) controls expression of alcohol oxidase which catalyzes the first step in methanol metabolism. To overcome the low specific activity of the enzyme, large quantities of alcohol oxidase are produced. Typically greater than 30% of the total soluble protein in methanol-induced cells is alcohol oxidase. The AOX1 promoter has been characterized and incorporated into a series of *Pichia* expression vectors in order to take advantage of the powerful AOX1 promoter to drive high-level expression of recombinant proteins.

P. pastoris is a useful candidate for expression of proteins in sufficient quantities (see the review in Romanos, M.A. et al. (1992)*Yeast* 8, 423-488). Protein yields of more than 1 gram per liter have been described (Tschopp, J.F. et al. (1987) 5,1305-1308; Paifer, E. et al. (1994)*Yeast* 10, 1415-1419; Laroche, Y. et al. (1994) *Bio/Technology* 12, 1119-1124; and Rodriguez, M. et al. (1994) *J. Biotechnol.* 33, 135-146). Though every protein will

express at a different level, most proteins are expressed at higher levels in *P. pastoris* than in bacterial, insect, or mammalian systems. *P. pastoris* is easily adaptable to large-scale fermentation for the production of recombinant protein. Expression of recombinant proteins in *P. pastoris* has been carried out in fermentors as small as 1 liter and as large as 10,000 liters. Moreover, recombinant protein expression in *P. pastoris* is less expensive than expression in insect or mammalian systems. The growth and expression medium does not require expensive supplements such as fetal bovine serum. In addition, expensive equipment such as CO_2 incubators and tissue culture hoods are not required. Because there is a long history, and therefore a wealth of accumulated knowledge, regarding the growth of yeast cells in large fermentors, such as is used in the brewing of beer or the production of penicillin, kilogram quantities of a protein can easily be purified.

One of the disadvantages of yeast expression systems in general is that recombinant proteins are frequently degraded at the N- and/or C-terminus (Romanos, M.A. et al. (1992)*Yeast* 8, 423-488; Scorer, C.A. et al. (1993) *Gene* 136, 111-119; Clare, J.J. et al. (1991) *Gene* 105, 205-212; Heim, J. et al. (1994) *Eur. J. Biochem.* 226, 341-353). This can cause problems in the functioning of the protein in some cases. For example, it has been shown that the C-terminal arginines of the anaphylatoxins C3a and C5a are necessary for full biological activity. Skidgel, R.A. (1988)*Trends Pharmacol. Sci. 9,* 299-304. Some protease deficient *P. pastoris* strains are available (Ohi, H. et al. (1996) Y. *Yeast* 12, 31-40).

What is needed in the art, therefore, are improved

recombinant constructs and methods of producing proteins in *P. pastoris* that do not result in undesired cleaved protein products.

Summary of the Invention

The present invention provides methods and compositions for recombinant expression of full length proteins in *P. pastoris*. The invention provides genetic constructs containing a disruption in the KEX1 gene that prevent cleavage of basic amino acids, such as lysine, from the carboxy terminal.

Therefore, it is an object of the present invention to provide improved recombinant products expressed in the organism *P. pastoris*.

It is an object of the present invention to provide recombinant protein products expressed in the organism *P. pastoris* that are not undesirably cleaved at the carboxy terminal.

It is an object of the present invention to provide improved recombinant constructs containing a disruption in the KEX1 gene of *P. pastoris* that prevents cleavage of basic amino acids, such as lysine, from the carboxy terminal.

It is an object of the present invention to provide methods of use for improved recombinant constructs containing a disruption in the KEX1 gene of *P. pastoris* that prevent cleavage of basic amino acids, such as lysine, from the carboxy terminal.

It is also an object of the present invention to provide improved recombinant constructs encoding a basic amino acid, such as lysine, at the carboxy terminal for use in a vector comprising a disrupted KEX1 gene of *P. pastoris* that prevent cleavage of basic amino acids, such as lysine, from the carboxy terminal.

These and other objects of the invention will become apparent to those skilled in the art upon a review of the detailed description and examples herein.

Brief Description of the Drawings

Figures 1A and 1B illustrate the sequence of the Kex1p enzyme from *P. pastoris*. Figure 1A is the amino acid sequence of *P. pastoris* Kex1p(SEQ ID NO:1). Arrows indicate amino acid residues likely to be important for the catalytic activity of Kex1p. Filled triangles denote potential N-linked glycosylation sites.

Figure 1B is a comparison of the amino acid sequences of S. *cerevisiae* (top)(SEQ ID NO:2) and *P. pastoris* Kex1p (bottom) (SEQ ID NO:3).

Figures 2A and 2B illustrates the genomic locus of the wild-type SMD1168 and the mutant kex1::SUC2 strains. Figure 2A is a genomic map around the wild type KEX1 gene. The numbers in parentheses correspond to the base pair number in the GenBank file (AF095574).

Figure 2B is a genomic map of the kex1::SUC2 disruption strain (ol is oligonucleotides used for verification of successful disruption of the KEX1 gene).

Detailed Description of the Invention

The present invention provides a recombinant nucleic acid construct comprising a disrupted KEX1 gene of *Pichia* that prevents cleavage of one or more basic amino acids from the carboxy terminal of a protein expressed therewith. In preferred embodiments, the *Pichia* is a *Pichia pastoris*. In other preferred

embodiments, the basic amino acid is a lysine. In further preferred embodiments, the disruption is a nucleic acid deletion, insertion or addition within the KEX1 gene.

The present invention also provides a method of expressing a full length protein comprising transfecting a gene encoding the protein into a recombinant nucleic acid construct and promoting the expression of the gene, wherein the recombinant nucleic acid construct comprises a disrupted KEX1 gene of *Pichia* that prevents cleavage of one or more basic amino acids from the carboxy terminal of a protein expressed therewith. In preferred embodiments, the *Pichia* is a *Pichia pastoris*. In other preferred embodiments, the basic amino acid is a lysine. In further preferred embodiments, the disruption is a nucleic acid deletion, insertion or addition within the KEX1 gene.

The invention also provides a method for expressing a full length protein comprising transfecting a gene encoding the protein into a recombinant nucleic acid construct and promoting the expression of the gene, wherein the recombinant nucleic acid construct comprises a disrupted KEX1 gene of *Pichia* that prevents cleavage of one or more basic amino acids from the carboxy terminal of a protein expressed therewith, and wherein the gene has been modified to contain one or more basic amino acids at the carboxy terminal. In preferred embodiments, the *Pichia* is a *Pichia pastoris*. In other preferred embodiments, the basic amino acid is a lysine. In further preferred embodiments, the disruption is a nucleic acid deletion, insertion, susbstitution or addition within the KEX1 gene.

By the use of "a" or "an" herein is meant or more than one, depending upon the context. By "protein" is meant any naturally occurring or modified sequence of amino acids, i.e. a polypeptide of any length. Such a protein can be a peptide fragment of a naturally occurring or modified protein, in addition to chimeric, fusion or labelled proteins, for example. The invention provides that a basic amino acid can be a lysine, histadine, or arginine. Under some physiological conditions, basic amino acids can include asparagine and glutamine. Cleavage of such basic residues can be prevented by the present invention whether the basic amino acid is at or near the carboxy terminal of the protein of interest. The invention contemplates that more than one basic amino acid may be located at or near the carboxy terminal of a protein in order to facilitate expression of an increased amount of the full length of the desired protein.

Accordingly, the present invention provides a KEX1 disruption strain wherein the disruption of the KEX1 reading frame allows expression of an increased amount of full length protein with the carboxy terminal intact. The KEX1 disruption strain is useful for the full length expression of other proteins with a carboxy terminal, or carboxy terminal region, basic amino acid. Examples of protein that can be expressed with the present invention include, but are not limited to, endostatin™ protein, and the anaphylatoxins C3a and C5a. It has been shown that the carboxy terminal arginines of C3a and C5a are necessary for full biological activity.

The 3.5 kb *P. pastoris* KEX1 gene locus has been deposited in the GenBank database and is available under the Accession

Number AF095574. Any disruption of the KEX1 gene that results in the production of full length protein as desired is contemplated by the present invention. One such example of a disruption involving a deletion and insertion is described below in the examples wherein a portion of the KEX1 gene is deleted and replaced with a portion of a *Sarcomyces cerevisiae* SUC2 gene.

The KEX1 disruption constructs of the present invention are also useful for producing proteins having a basic amino acid purposefully added to the carboxy terminus in order to protect the carboxy terminus from degradation by other carboxypeptidases. The present invention provides for the inhibition of degradation of recombinant proteins by genetically constructing the protein with a carboxy terminal region lysine, or other basic amino acid(s), that is compatible with protein activity.

In a related embodiment, the KEX1 strain can be used to make epitope labeled proteins. In recent years, epitope tagging of proteins has become useful, because it allows the detection of proteins with highly specific, high-affinity antibodies within several weeks without the need to express a sufficient quantity of protein for injection into rabbits and without the need to depend on adequate antibody titers. The FLAG-tag (Asp-Tyr-Lys-Asp_4-Lys)(SEQ ID NO:4) is among the frequently used amino acid sequences. Expression of secreted proteins tagged on the carboxy terminus with a FLAG-tag in *P. pastoris* results in the degradation of the carboxy terminal lysine and the protein tag will not be recognized anymore by the FLAG antibody. The KEX1 disruption of the present invention avoids this degradation.

Therefore, the present invention provides methods and genetic constructs which increase the amount of full length protein expressed therewith. In some cases with expression of protein on the KEX1 disrupted construct, a mixture of proteins is formed, some having intact carboxy terminii and some not intact. It appears that lysis of yeast cells occurs which releases proteases into the culture-supernatant. The invention provides a means to overcome this problem by using a double deletion strain, such as crossing the KEX1 disruption strain with the PRC1 deletion strain (Ohi, H. et al. (1996) *Yeast* **12,** 31-40). Carboxy terminal degradation can also be minimized by a short methanol induction phase. In the KEX1 deletion strain, degradation of the carboxy terminal lysine by other enzymes can be minimized using either shaker flask purification or short fermentation runs. Furthermore, intact proteins can easily be separated from degraded forms using cation exchange chromatography. Thus, the invention provides that the disrupted KEX1 constructs allow for an increased amount of full length protein to be expressed therein.

Furthermore, other embodiments of the invention provide that the cloning of the disrupted KEX1 gene allows construction of an overexpression strain, which may be more efficient in the cleavage of carboxy terminal basic amino acids. This strain is useful during expression of proteins, where proteolytic activation and removal of carboxy terminal basic amino acids is important for the generation of active protein. A classical example of such a protein is insulin. Construction of a *P. pastoris* strain overexpressing the Kex1p, Kex2p and Ste13p proteins is very

efficient in secreting active insulin. Thim et al. (*Proc. Natl. Acad. Sci.* 83, 6766-6770 (1986)) have reported that in *S. cerevisiae* the proper processing of insulin is a limiting factor.

In the examples, the preferred SMD1168 strain of *P. pastoris* was used. SMD1138 is a pep4 mutant strain of *P. pastoris* deficient in proteinase A activity. Proteinase A is required for the proteolytic activation of a number of proteases, including carboxypeptidase Y. SMD1168 can be used to reduce proteolysis of some recombinant proteins expressed in the strain.

In the examples, the preferred pPIC9K *Pichia* expression vector, also available from Invitrogen, was used. The pPIC9K *Pichia* expression vector allows one to obtain *Pichia* strains that contain multiple copies of the gene of interest. Several publications suggest that increasing the number of copies of the gene of interest in a recombinant *Pichia* strain may increase protein expression levels. The pPIC9K vector carries the kanamycin resistance gene which confers resistance to G418 in *Pichia*. Spontaneous generation of multiple insertion events, which occur in *Pichia* at a frequency of 1-10%, can be identified by the level of resistance to G418. *Pichia* transformants are selected on histidine deficient medium and screened for their level of resistance to G418. An increased level of resistance to G418 indicates multiple copies of the kanamycin resistance gene and of the gene of interest. The pPIC9K vector allows secreted expression of the gene of interest.

The pAO815 vector is specially designed to generate multiple copies of the gene of interest in a single vector. The vector contains a Bgl II site upstream of the 5' AOX1 gene and a unique BamH I

site downstream of the 3' AOX1 transcription termination (TT) signal. Four steps are performed to generate multiple copies of the gene of interest. First, the gene is cloned into the unique EcoR I site in the vector. Then, the construct is digested with BamH I and Bgl II to release the "expression cassette" containing the AOX1 promoter, gene of interest, and 3' AOX1 TT. Next, multiple copies of the expression cassette are generated *in vitro* by ligation. Finally, the multiple copies are inserted back into the pAO815 vector and transformed into *Pichia.*

It should be understood that the compositions and methods disclosed and claimed herein are useful in other species and strains of *Pichia* and that other expression systems can be used. For example, Invitrogen also offers a *Pichia methanolica* expression system, to which the presently disclosed methods are applicable.

The invention is further illustrated by the following examples, which are not to be construed in any way as imposing limitations upon the scope thereof. On the contrary, it is to be clearly understood that resort may be had to various other embodiments, modifications, and equivalents thereof, which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the present invention.

Examples

Endostatin™ protein is a potent angiogenesis inhibitor. When endostatin™ protein is expressed in *Pichia pastoris*, analysis of the expressed protein by mass spectrometry indicates that the protein is truncated. N-terminal sequence analysis determined that the N-

terminus was intact, suggesting that the C-terminal lysine was missing. In *Saccharomyces cerevisiae,* Kex1p, a carboxypeptidase B-like enzyme, can cleave lysine and arginine residues from the C-terminus of peptides and proteins. It was discovered that the KEX1 homologue in *P. pastoris* is responsible for the loss of the C-terminal lysine of endostatin™ protein.

Example 1: Construction of a *P. pastoris* genomic library

The polymerase chain reaction was performed according to standard protocols using either TAQ (Boehringer Mannheim) or PFU polymerase (Stratagene). All PCR fragments were cloned into the plasmid pCR™2.1, a TA-cloning vector (Invitrogen). PFU amplified fragments were isolated with the PCR-QIA Quick spin purification kit (Qiagen), incubated for 10 min with TAQ polymerase and ligated into the TA-vector. PFU polymerase was used for the amplification of the KEX1 gene locus from genomic DNA of the *P. pastoris* strain SMD1168 and the SUC2 gene from the *S. cerevisiae* strain S288C (ATCC 204508). Sequence analysis was performed on an ABI Prism Model 377, Version 3.0, in only one direction.

For PCR analysis of the KEX1 disruption strains, single colonies were inoculated into 1.5 ml YEPD medium and grown overnight to saturation. The cell pellet was resuspended in 200 µl of lysis buffer (1% SDS, 2% Triton X-100, 100 mM NaCl, 10 mM Tris pH 8.0, 1 mM EDTA). After resuspension, 200 µl of phenol-chloroform and 300 mg of acid-washed glass beads were added and the solution was vortexed for 2 minutes. Following a 5 min centrifugation step, the supernatant was ethanol precipitated and

1/50 was used for PCR analysis.

DNA was isolated from the *P. pastoris* strain SMD1168 using the Qiagen Genomic DNA Purification Kit. 2 μg of genomic DNA were digested for 4 h with 30 Units *Eco*R1. The digested DNA was repurified with the QIAEX II kit and 100 ng were ligated into the Lambda ZAP II vector according to the protocol supplied by the manufacturer (Stratagene). 3 μl of the ligation mix were used for packaging the vector into phage particles (Gigapack III Gold packaging kit, Stratagene). Eleven million primary clones were obtained and 200,000 clones were amplified. The Qiagen Lambda DNA isolation protocol was used to obtain phage DNA from agarose plate lysates. 200 ng of phage DNA were used for PCR analysis.

Example 2: Expression of endostatin™ protein from *P. pastoris*

The murine and human endostatin™ protein cDNAs were amplified using PCR and cloned into the *Xho*1/*Not*1 sites of the expressing vector pPIC9 (pPIC9-mES and pPIC9-hES, respectively, Invitrogen). This cloning created a fusion protein of the alpha-factor secretion signal with endostatin. Murine endostatin™ protein should be secreted starting with the published N-terminus His-Thr-His-Gln-Glu-Phe and human endostatin™ protein should start with His-Ser-His-Arg-Glu-Phe, if correct processing by the Kex2p and the Ste13p occurs. These constructs were verified by sequencing. A 1.6 kb *Bam*H1/*Xba*1 fragment of pPIC9-mES and pPIC9-hES was subcloned into pPIC9K.

The plasmid was transfected into the SMD1168 strain. This

strain has a deletion in the PEP4 gene, which is important for the activation of carboxypeptidase Y (CPYp). Although the CPYp is not the only carboxypeptidase in yeast cells, the SMD1168 strain should have a lower capacity to degrade proteins from the C-terminus compared to the wild type *P. pastoris* strain GST115. Integration of the murine construct into SMD1168, followed by selection on -His and G-418 plates were performed according to the manufacturer's instructions (*Pichia* Expression Kit, Invitrogen).

Example 3: Purification of endostatin™ protein from *P. pastoris*

Murine endostatin™ protein was purified from yeast supernatant grown in shaker flasks. For shaker flask purification, the recommendations of the manufacturer were followed (Invitrogen). Rich medium buffered with potassium phosphate at pH 6.0 was used. The cultures were induced with methanol for 48-72 h. During the induction period the pH increased to 6.5. Cells were removed by centrifugation and the supernatant was slowly mixed with a saturated ammonium sulfate solution to a final concentration of 45% ammonium sulfate. The ammonium sulfate solution was prepared by dissolving 767 g of ammonium sulfate (ICN) in 1 L of 10 mM Tris, pH 7.0. The pH was readjusted to pH 7.0 and the solution stored at 4° C. The ammonium sulfate/supernatant solution was stirred for 2 h at 4° C and centrifuged at 5000 rpm for 20 min. The precipitated proteins were resuspended in 0.1-0.2 starting volumes of phosphate buffered saline (PBS). The solution was dialyzed against PBS and directly applied to a heparin-Sepharose column. After washing the column

with PBS and 0.25 M NaCl, 20 mM Tris, pH 7.4, endostatin™ protein was eluted using a 0.25 to 1.5 NaCl gradient. Fractions containing endostatin™ protein were pooled, dialyzed against PBS and filter sterilized. The protein was stored at -20° C.

Murine and human endostatin™ protein were also purified from yeast culture-supernatant harvested from a fermenter (5 liter bench-top B. Braun fermenter). Complex medium was used and the fermentation conditions recommended by Invitrogen were followed. The pH was kept constant at 6.5. The pH of the complex medium was kept constant at 6.5, because it has been shown that endostatin™ protein is a zinc-binding protein. In addition, zinc-binding may be necessary for full biological activity and it protects the N-terminal amino acids from proteases and therefore endostatin™ protein from inactivation (Boehm et al., submitted). Growing the yeast cells below pH 6.5 may cause dissociation of zinc from the binding pocket and may allow proteases to cleave off some N-terminal amino acids, which would result in the purification of inactive protein. Atomic absorption of endostatin™ protein isolated from shaker flasks and a fermenter showed a 1:1 ratio of zinc to protein.

The methanol induction phase was either 40 h or 70 h. The supernatant was dialyzed against PBS and concentrated. Ammonium sulfate precipitation was not possible and the solution was directly applied onto a heparin-Sepharose column. After washing the column with PBS and 0.25 M NaCl, 20 mM Tris, pH 7.4, endostatin™ protein was eluted using a 0.25 to 1.5 M NaCl gradient. Fractions containing endostatin™ protein were pooled and

dialyzed against PBS. A second chromatography step was necessary to purify endostatin™ protein to apparent homogeneity. A HiPrep 26/60 Sephacryl 3-200 HR gel filtration column (Pharmacia) was used, equilibrated in PBS. Endostatin™ protein containing fractions were filter sterilized and stored at -20° C at a protein concentration of 0.5-2.0 mg/ml, determined using the Bio-Rad Bradford assay with immunoglobulin as the standard protein.

Endostatin™ protein was purified to apparent homogeneity from the shaker flasks as shown by SDS PAGE. N-terminal sequence analysis showed the correct N-terminus but mass spectrometry revealed that one amino acid is missing from the C-terminus. The molecular weight determined by mass spectrometry was 20248.4 Da, which is in good agreement with the calculated molecular weight of endostatin™ protein without a C-terminal lysine of 20243.79. The "real" calculated molecular weight of endostatin™ protein without the C-terminal lysine would be 20247.79 but 4 Da have been subtracted because two disulfide bonds are formed. Free cysteines in endostatin™ protein purified from the *P. pastoris* culture medium were not found. The formation of two disulfide bonds was also confirmed by the two recently published structures of mouse and human endostatin™ protein (Hohenester, E. et al., (1998) *EMBO J.* 17, 1656-1664; Ding, H-Y. et al. (1998)*Proc. Natl. Acad. Sci.* 95, 10443-10448). Mass spectrometry analysis also demonstrated that endostatin™ protein is not glycosylated when expressed in *P. pastoris*.

Expression of endostatin™ protein in a fermenter required the use of a heparin-Sepharose and a gel-filtration step to purify it

to apparent homogeneity. Under the high cell density growth conditions during fermentation a fraction of endostatin™ protein was further degraded. Mass spectrometry analysis indicated that the C-terminal serine exposed by removal of the lysine was removed. The calculated molecular weight of endostatin™ protein without a C-terminal lysine and serine and without 4 Da due to the formed disulfide bonds is 20156.7 Da.

Example 4: Cloning of the KEX1 gene

In *S. cerevisiae*, the KEX1 gene encodes a carboxypeptidase B-like enzyme, which is involved in the processing of the killer toxin and the alpha-factor (Dmochowska, A. et al. (1987) *Cell* 50, 573-584). The Kex2p and the Ste13p cleave off the alpha-factor signal sequence, which allows secretion of alpha-factor or the protein of interest into the culture medium. Because the alpha-factor secretion signal was used for expression of endostatin™ protein, it was hypothesized that the Kex1p homologue in *P. pastoris* removed the C-terminal lysine. The *P. pastoris* KEX1 gene was cloned and disrupted in order to test this hypothesis.

A PCR approach was used to clone the KEX1 gene. The S. *cerevisiae* Kex1p and the CPYp from *S. cerevisiae* and *P. pastoris* contain a highly conserved motif around the active serine residue, Gly-Glu-Gly-Ser-Tyr-Ala-Gly (SEQ ID NO:5). Degenerative oligonucleotides of this motif were designed and as a second primer T3 from the vector sequence was used. 200 ng of phage library DNA were used as a template. The following degenerate oligonucleotide combined with a T3 primer from the lambda vector (pBluescript) amplified a 1.8 kb fragment at 62° C annealing

temperature, 5'-TG NCC CGC RTA RCT YTC ACC-3'(SEQ ID NO:6). Sequence analysis followed by a BLAST database search showed the highest homology to the S. *cerevisiae* KEX1 gene. The T7 primer from pBluescript and a reverse primer obtained from the sequence of the 1.8 kb PCR fragment were used to obtain a 4.2 kb fragment 3' of the Gly-Glu-Gly-Ser-Tyr-Ala-Gly (SEQ ID NO:5) motif. A total of 3.5 kb were sequenced. Sequence and protein analysis were done with the Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, Wisc. In the KEX1 gene there are two ATG codons in a row. The second ATG was used as the first amino acid codon. The reading frame encodes for 623 amino acids with a predicted molecular weight of 70,017.26 Da (Fig. 1A). The *S. cerevisiae* KEX1 gene encodes for 729 amino acids. The difference of 106 amino acids is mainly due to a shortened C-terminus. The *P. pastoris* Kex1p is 75 amino acids shorter on the C-terminus compared to the *S. cerevisiae* homologue. Amino acid comparison shows that these two proteins are only 36% identical and 43.7% similar (Fig. 1B). The N- and C-terminal regions are especially weakly conserved. Comparison of the two proteins without the first 80 and last 160 amino acids of the *P. pastoris* Kex1p increased the identity to 48% and the similarity to 57%. Nevertheless, residues belonging to the catalytic triad of the *S. cerevisiae* Kex1p (Ser198, Asp406 and His470) are conserved (Shilton, B.H. et al. (1997) *Biochemistry* 36, 9002-9012). In addition, the amino acid residues close to these three residues are highly conserved. The glutamic acids (Glu110 and Glu197 in the *S. cerevisiae* Kex1p) that are important for the hydrolysis of peptides

are also conserved in the *P. pastoris* Kex1p (Glu99 and Glu176, Stennicke, H.R. et al. (1996) J. *Biochemistry* 35, 7131 7141). Furthermore, the hydropathy profile determined according to Engelman, D.M. et al. (1986) *Annu. Rev. Biophys. Chem.* 15, 321-353 and Kyte, J. et al. (1982) *J. Mol. Biol.* 157, 105-132 showed a very similar profile indicating structural conservation (not shown).

A significant difference between these two sequences is that the *P. pastoris* Kex1p does not have the very distinct 105 residue stretch rich of aspartic and glutamic acids found in the *S. cerevisiae* sequence (Fig. 1B). Nevertheless, the *P. pastoris* Kex1p has a short hydrophilic-acidic region upstream of the conserved membrane spanning domain. This transmembrane domain might target the Kex1p to the late Golgi compartment, as has been shown for the *S. cerevisiae* Kex1p (Cooper, A. et al. (1989) *Mol. Cell. Biol.* 9, 2706-2714; Bryant, N.J. et al. (1993) *J. Cell Science* 106, 815-822). A second hydrophobic region is located at the N-terminus and may serve as a signal sequence (Watson, M.E.E. (1984) *Nucl. Acids Res.* 12, 5145-5156).

The *P. pastoris* Kex1p has six potential N-linked glycosylation sites (AspXxxSer/Thr) compared to four in the *S. cerevisiae* homologue (Fig. 1A). The two sites between amino acid 449-469 in the *S. cerevisiae* Kex1p and 430-440 in *P. pastoris* Kex1p are conserved (Fig. 1B).

Example 5: Disruption of the KEX1 gene with SUC2

In the next step major parts of the reading frame were disrupted and the question of whether endostatin™ protein could be purified with the C-terminal lysine was investigated. In order to

disrupt the KEX1 gene, a strategy was followed using the SUC2 gene from *S. cerevisiae* as a selectable marker (Figs. 2A and 2B). The SUC2 locus encodes the invertase gene, which should allow *P. pastoris* cells to grow efficiently on minimal sucrose plates. *P. pastoris* cells grow slowly on minimal sucrose (MS) plates.

A 2.2 kb fragment of the KEX1 locus including the ATG and the stop codon was amplified from SMD1168 genomic DNA (Fig. 2B) and cloned into the TA-cloning vector pCR™2.1. The oligonucleotides contained a 5' *Sma*1 and 3' *Sna*B1 restriction site. The resulting plasmid (pCR™2.1-KEX1) was amplified in the methylation deficient strain SCS110 (Stratagene) and cut with *Cla*1/*Nde*1. A 2.9 kb segment of the SUC2 gene was PCR amplified from 5288C genomic DNA with *Cla*1 and *Nde*1 flanking restriction sites and cloned into the pCR™2.1-KEX1 plasmid. The disruption plasmid (pCR™2.1-kex1::SUC2) was digested with *Sma*1/*Sna*B1, transformed into SMD1168 using the Frozen-EZ Yeast Transformation II kit (Zymo Research) and the yeast cells were plated on minimum sucrose plates prepared as described by Ohi, H. et al. (1996) *Yeast* 12, 31-40. Positive colonies were restreaked on MS plates and analyzed by PCR. Successful disruption of the KEX1 gene was detected using three different PCR primer pairs.

One of the SMD1168 kex1::SUC2 strains was transformed with pPIC9K-mES and pPIC9K-hES and selected on -HIS and consequently on G-418 plates. The endostatin™ protein expression vectors were transformed into the KEX1 deletion strain. Expression and purification of murine and human endostatin™

protein was performed essentially as described above.

Example 6: Mass spectrometry analysis

Recombinant endostatin™ protein was assayed on a Voyager DE-Elite or Voyager DE-STR Biospectrometry™ workstation (Perceptive Biosystems). Molecular mass analysis was performed on 2 µl purified protein in PBS cocrystallized with 2 µl of a 10 mg/ml sinapinic acid/myoglobin solution (3,5-dimethoxy-4-hydroxy-cinnamic acid in 1:2 v/v acetonitrile/water, 0.1 % TFA and 200 ng/ml horse myoglobin (Sigma) as internal standard). Under these conditions the error of the mass determination is below 0.1%. In the case of endostatin, this gives an expected error of less than 20 Da. The determined masses deviate by less than 10 Da from the calculated values. Therefore, single amino acid changes can easily be detected.

Mass spectrometry analysis of murine endostatin™ protein showed a molecular weight of 20378.7 Da, in agreement with the calculated weight of 20371.79. Four Da have been subtracted due to the formed disulfide bonds. This result shows that disruption of the KEX1 gene in *P. pastoris* allows expression of full length endostatin™ protein. Fermentation for 40 h resulted in the purification of intact endostatin™ protein. However, after a 70 h fermentation run mass spectrometry analysis of purified endostatin™ protein revealed that it was again degraded. A signal corresponding to endostatin™ protein without the C-terminal lysine appeared as well as a small peak without both the C-terminal lysine and serine.

Example 7: Cation exchange chromatography

The isoelectric point of intact murine endostatin™ protein is 8.93. Removal of the C-terminal lysine shifts the isoelectric point to 8.54. This pH difference was exploited to separate intact from degraded endostatin™ protein using cation exchange chromatography. A MonoS HR 5/5 column (Pharmacia) was used to separate intact endostatin™ protein and endostatin™ protein without the C-terminal lysine. 0.1 to 1 mg of purified endostatin™ protein in PBS was loaded onto the column using the FPLC system from Pharmacia. The column was washed with 50 mM HEPES, pH 8.0 and endostatin™ protein was eluted with a 0 to 0.35 M NaCl gradient.

Two protein peaks eluted from the column. The first peak (approximately eluting at 180 mM NaCl) corresponds to endostatin™ protein without a C-terminal lysine and to a certain extent without serine, whereas the second peak (approximately eluting at 220 mM NaCl) was intact endostatin™ protein as determined by mass spectrometry. The percentage ratio of intact to degraded endostatin™ protein was about 52 to 48.

Human endostatin™ protein was also expressed in the KEX1 gene deletion strain. Most of the protein showed a molecular weight close to the calculated mass of 20091.5 Da (4 Da have been subtracted). There was also a signal corresponding to human endostatin™ protein without a C-terminal lysine. The isoelectric point of human endostatin™ protein drops from 9.49 to 8.93 without the C-terminal lysine. The degraded form could be separated from the intact form on a MonoS column. 87% of human

endostatin™ protein was full length. The calculated mass of human endostatin™ protein without the lysine is 19663.4 Da.

The above description is intended to be illustrative and not restrictive. Many embodiments will be apparent to those of skill in the art upon reading the above description. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. The disclosures of all articles and references referred to herein, including patents, patent applications, and publications, are incorporated herein by reference.

What is claimed is:

1. A recombinant nucleic acid construct comprising a disrupted KEX1 gene of *Pichia* that prevents cleavage of one or more basic amino acids from the carboxy terminal of a protein expressed therewith.

2. The construct of Claim 1, wherein the *Pichia* is a *Pichia pastoris*.

3. The construct of Claim 1, wherein the basic amino acid is a lysine, arginine or histadine.

4. The construct of Claim 1, wherein the basic amino acid is a lysine.

5. The construct of Claim 1, wherein the disruption is a nucleic acid deletion and insertion within the KEX1 gene.

6. The construct of Claim 1, wherein the disruption is a nucleic acid deletion within a portion of the KEX1 gene, and replacement thereof with a portion of *Sarcomyces cerevisiae* SUC2 gene.

7. A method for expressing a full length protein comprising transfecting a gene of interest into a recombinant nucleic acid construct and promoting the expression of the gene of interest, wherein the recombinant nucleic acid construct comprises a disrupted KEX1 gene of *Pichia* that prevents cleavage of one or more basic amino acids from the carboxy terminal of a protein expressed therewith.

8. The method of Claim 7, wherein the *Pichia* is a *Pichia pastoris*.

9. The method of Claim 7, wherein the basic amino acid is a lysine, arginine or histadine.

10. The method of Claim 7, wherein the basic amino acid is a lysine.

11. The method of Claim 7, wherein the disruption is a nucleic acid deletion and insertion within the KEX1 gene.

12. The method of Claim 7, wherein the disruption is a nucleic acid deletion within a portion of the KEX1 gene, and replacement thereof with a portion of *Sarcomyces cerevisiae* SUC2 gene.

13. A method for expressing a full length protein comprising transfecting a gene of interest into a recombinant nucleic acid construct and promoting the expression of the gene encoding the protein, wherein the recombinant nucleic acid construct comprises a disrupted KEX1 gene of *Pichia* that prevents cleavage of one or more basic amino acids from the carboxy terminal of a protein expressed therewith, and wherein the gene has been modified to contain one or more basic amino acids at the carboxy terminal of the protein.

14. The method of Claim 13, wherein the *Pichia* is a *Pichia pastoris*.

15. The method of Claim 13, wherein the basic amino acid is a lysine, arginine or histadine.

16. The method of Claim 13, wherein the basic amino acid is a lysine.

17. The method of Claim 13, wherein the disruption is a nucleic acid deletion and insertion within the KEX1 gene.

18. The method of Claim 13, wherein the disruption is a nucleic acid deletion within a portion of the KEX1 gene, and replacement thereof with a portion of *Sarcomyces cerevisiae* SUC2 gene.

19. A protein expressed by the method of Claim 7.

20. A protein expressed by the method of Claim 13.

1 MLKLLCLLLP LVAVSASPID LGSQKDYLVL DLPGLSHLSE TQRPTMHAGL

51 LPLNLSFVAD DDTEYFFWRF SKQDVDRADI VFWLNGGPGC SSMDGALMEL

101 GPFVINPKQE VEYNEGTWVE AADVVFVDQP GGTGFSSTTN YLTELTEVAD

151 GFVTFLARYF HLFPADVYKK FTLGGESYAG QYVPYILKAI MDDLKSDSGQ

201 LPKELYLKGA LIGNGWIDPN EQSLSYLEFF IKKELIDIHG SYMPGLLQQQ

251 EKCQNLINHS SGEASESQIS YSACEKILND ALRFTRDKKA PLDQQCINMY

301 DYTLRDTYPS CGMSWPPYLP DVTAFLQKKS VLEALHLDSS ASWSECSARV

351 GSHLKNKISV PSVDILPDLL QEIPIILFNG DHDIICNCIG TERMIDKLEF

401 NGDQGFTEGT EYIPWFYNEV NVGKVISERN LTYVRVYNSS HMVPFDNTPV

451 SRGLLDIYFD NFEDVEYNNV SGIATPVYDV DKNITYIDSN DPRLQNGPKS

501 SSTDDSAAHG NPFFYYVFEL FVIVLLLCGL VYLYQYYSNS APHSILADKH

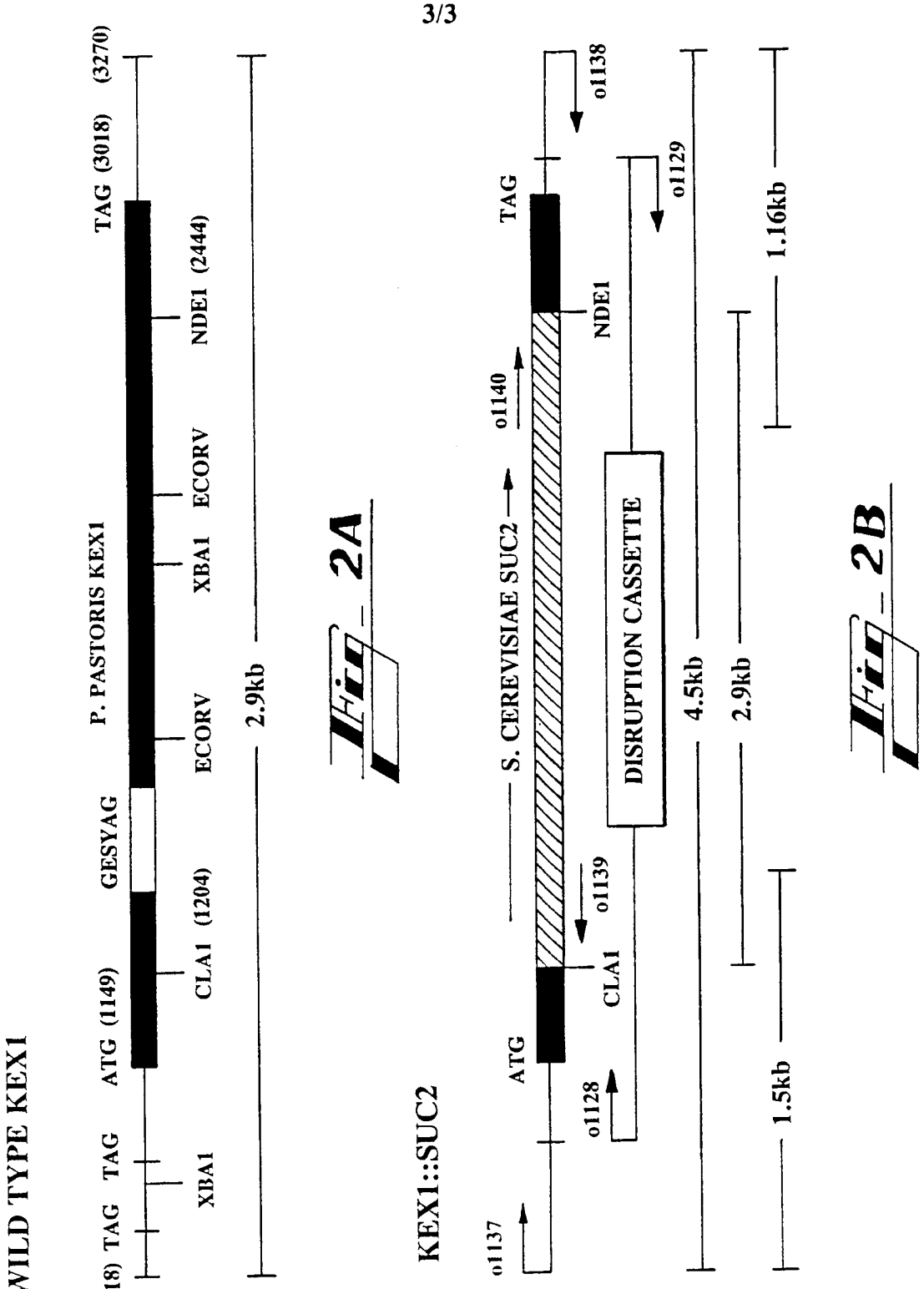
551 KKKSKNKSKN VRFLDDLESN LDLDNTDDKK DNSVMSKLLS SMGYQAQEPY

601 KPLDKGANAD LDIEMDSHGT SEK*

Figure 1A

```
  1 MFYNRWLGTWLAMSALIRISVSLPSSEEYKVAYELLPGLSEVPDPSNIPQ 50
    :  |   | : |.   : | | .:| |    |||||. . :   |
  1 ..MLKLLCLLLPLVAVSASPIDLGSQKDYLVLD..LPGLSHLSETQR.PT 45

 51 MHAGHIPLRSEDADEQDSSDLEYFFWKFTNNDSNGNVDRPLIIWLNGGPG 100
    ||||  :||      : |.   |||||:|.  | .      :: |||||||
 46 MHAGLLPLNLSFVADDDT...EYFFWRPSKQDVD...RADIVFWLNGGPG 89

101 CSSMDGALVESGPFRVNSDGKLYLNEGSWISKGDLLFIDQPTGTGFSVEQ 150
    |||||||||.|  |||  :|   ..  |||.|:   |..|:|||  |||||
 90 CSSMDGALMELGPFVINPKQEVEYNEGTWVEAADVVFVDQPGGTGFSSTT 139

151 NKDEGKIDKNKFDEDLEDVTKHFMDFLENYFKIFPEDLTRKIILSGESYA 200
    |          :  :| :|   |. ||  || :|| |. :|  | |||||
140 N..........YLTELTEVADGFVTFLARYFHLFPADVYKKFTLGGESYA 179

201 GQYIPFFANAILNHNKFSKID.GDTYDLKALLIGNGWIDPNTQSLSYLPF 249
    |||:|:   ||:.  |              ||  |||||||||| ||||||  |
180 GQYVPYILKAIMDDLKSDSGQLPKELYLKGALIGNGWIDPNEQSLSYLEF 229

250 AMEKKLIDESNPNFKHLTNAHENCQNLINSASTDEA.AHFSYQECENILN 298
   ..|.|||          |    | |||||| .| : . .  ||  || |||
230 FIKKELIDIHGSYMPGLLQQQEKCQNLINHSSGEASESQISYSACEKILN 279

299 LLLSYTRESSQKGTADCLNMYNFNLKDSYPSCGMNWPKDISFVSKFFSTP 348
    | :||:         |:|||.: |:|.||||||.||   :  |. |
280 DALRFTRDKKAPLDQQCINMYDYTLRDTYPSCGMSWPPYLPDVTAFLQKK 329

349 GVIDSLHLDSDKIDHWKECTNSVGTKLSNPISKPSIHLLPGLLESGIEIV 398
    |:::.|||||      | ||.  ||. | | || ||: :|| ||:  | |:
330 SVLEALHLDSSA..SWSECSARVGSHLKNKISVPSVDILPDLLQE.IPII 376

399 LFNGDKDLICNNKGVLDTIDNLKWGGIKGFSDDAVSFDWIHKSKSTDDSE 448
    |||||  |:|||  |    || |.. | .||.:      | :       .|
377 LFNGDHDIICNCIGTERMIDKLEFNGDQGFTEGTEYIPWFY.......NE 419

449 EFSGYVKYDRNLTFVSVYNASHMVPFDKSLVSRGIVDIYSNDVMIIDNNG 498
    | |  :||||:| |||.|||||||. . ||||:.|||  ..   :: |
420 VNVGKVISERNLTYVRVYNSSHMVPFDNTPVSRGLLDIYFDNFEDVEYNN 469

499 KNVMITTDDDSDQDATTESGDKPKENLEEEEQEAQNEEGKEKEGNKDKDG 548
  . . |   | |.. |    . |:
470 VSGIATPVYDVDKNITYIDSNDPRLQNGPKSSSTDDSAAHGNPFFYYVFE 519

549 DDDNDNDDDDEDDHNSEGDDDDDDDDDEDDNNEKQSNQGLEDSRHKSSEY 598
                                  | . .|  |.          |
520 LFVIVLLLCGLVYLYQYYSNSAPHSILADKHKKKSKNKSKNVRPLDDLES 569

599 EQEEEEVEEPAEEISMYKHKAVVVTIVTFLIVVLGVYAYDRRVRRKARHT 648
    : :  ::  :   | |  . .            |   |           |
570 NLDLDNTDDKKDNSVMSKLLSSMGYQAQEPYKPLDKGANADLDIEMDSHG 619

649 ILVDPNNRQHDSPNKTVSWADDLESGLGAEDDLEQDEQLEGGAPISSTSN 698

620 TSEK*..................................................  624
```

Figure 1B


WILD TYPE KEX1
(318) TAG
TAG
XBA1
ATG (1149)
CLA1 (1204)
GESYAG
ECORV
P. PASTORIS KEX1
XBA1
ECORV
NDE1 (2444)
TAG (3018)
(3270)
2.9kb
Fig. 2A
KEX1::SUC2
o1137
ATG
o1128
CLA1
o1139
S. CEREVISIAE SUC2
o1140
DISRUPTION CASSETTE
NDE1
TAG
o1129
o1138
1.5kb
2.9kb
4.5kb
1.16kb
Fig. 2B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/23351

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12N 15/81; C12P 21/02
US CL : 435/69.1, 320.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 320.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BRS-WEST, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,691,183 A (FRANZUSOFF et al.) 25 November 1997, col. 24, lines 25-50.	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 03 JANUARY 2000	Date of mailing of the international search report 21 JAN 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JAMES KETTER Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*